FEBS Letters 579 (2005) 302–312 FEBS 29171

Hypothesis

An elaboration on the *syn–anti* proton donor concept of glycoside hydrolases: electrostatic stabilisation of the transition state as a general strategy

W. Nerinckx*, T. Desmet, K. Piens[1], M. Claeyssens

Laboratory for Glycobiology, Department of Biochemistry, Physiology and Microbiology, Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

Received 29 October 2004; revised 10 December 2004; accepted 10 December 2004

Available online 21 December 2004

Edited by Stuart Ferguson

Abstract An in silico survey of all known 3D-structures of glycoside hydrolases that contain a ligand in the −1 subsite is presented. A recurrent crucial positioning of active site residues indicates a common general strategy for electrostatic stabilisation directed to the carbohydrate's ring-oxygen at the transition state. This is substantially different depending on whether the enzyme's proton donor is *syn* or *anti* positioned versus the substrate. A comprehensive list of enzymes belonging to 42 different families is given and selected examples are described. An implication for an early evolution scenario of glycoside hydrolases is discussed.

Keywords: Glycoside hydrolase; Substrate/ligand complex; Enzyme mechanism

1. Introduction

Almost 70 years ago, it was pointed out by Linus Pauling that the transition state theory is applicable to enzymes and that the catalytic rate constant depends mainly on the free energy of activation; enzyme catalysed reactions therefore involve stabilisation of the corresponding transition states [1]. Evolutionary pressure has equipped enzymes with an intricate microenvironment within their active site designed for this purpose. Many different factors (strain, desolvation, dynamic effects, entropic factors, and enthalpic factors such as hydrogen bonds, hydrophobic and electrostatic interactions) may – or may not – contribute to the affinity for the short-living transition state. In this respect, an excellent evaluation can be found in the work of Arieh Warshel, who demonstrated that "a complimentarity between the electrostatic potential of the enzyme-active site and the change in charges during the reaction" is the most significant catalytically contributing factor [2–4].

2. Mechanism of glycoside hydrolases

O-Glycoside hydrolases (EC 3.2.1.x), a structurally diverse group of multi-subsite enzymes, are abundantly present in Nature. On the basis of amino acid sequence similarity, they have hitherto been classified into more than 90 different families [5,6], and observation of fold conservation allowed some families to be grouped in clans [7]. With almost no exceptions, *O*-glycoside hydrolases possess at least two mechanistically important carboxyl functions at the junction of subsites −1/+1 [8]: (i) a lateral *syn* or *anti* positioned [9] proton donor close to the glycosidic oxygen; and (ii) a nucleophilic carboxylate residue (retaining hydrolases) or a solvent-nucleophile assisting carboxylate residue (inverting hydrolases). The general mechanism of a β-D-glucoside hydrolysis by retaining and inverting enzymes is represented in Fig. 1.

Hundreds of 3D-structures of *O*-glycoside hydrolases have already been solved, but for about half of all glycoside hydrolase (GH) families structural data are still missing. The availability of enzyme structures containing a ligand within the active centre, especially with the crucial −1 subsite occupied, is important for closer insight into ligand recognition, substrate specificity and modus operandi. This has been addressed for representative enzymes belonging to about 40 *O*-glycoside hydrolase families, by the use of specific inhibitors, cryo-techniques or site-directed mutagenesis. A major breakthrough was the development of 2-deoxy-2-fluoro-derivatives, which has allowed the covalent trapping of the glycosyl-enzyme intermediate of many retaining hydrolases [10]. The current understanding of glycosidase mechanisms has recently been reviewed thoroughly [11,12].

Although (poly)saccharide *O*-glycosidic bonds are remarkably stable towards uncatalysed hydrolysis at room temperature, with k_{uncat} values in the order of 10^{-15} s^{-1}, most of the glycoside hydrolases attain k_{cat} values in the order of 100 s^{-1} or higher, truly a stunning rate enhancement for enzymes which in most cases do not need assistance by either metals or cofactors. This analysis by Richard Wolfenden [13] further implies that the dissociation constant of a glycoside hydrolase-transition state complex ($ES^{\ddagger}$-K_{diss}) is in the order of 10^{-22} M, an extremely low value. It has two important corollaries:

*Corresponding author. Fax: +32 9 264 5332.
E-mail addresses: wim.nerinckx@ugent.be (W. Nerinckx), t.desmet@ugent.be (T. Desmet), kathleen@biotech.kth.se (K. Piens), marc.claeyssens@ugent.be (M. Claeyssens).

[1] Present address: Laboratory of Wood Biotechnology, Royal Institute of Technology, AlbaNova University Centre, KTH Biotechnology, SE-106 91 Stockholm, Sweden.

0014-5793/$22.00
doi:10.1016/j.febslet.2004.12.021

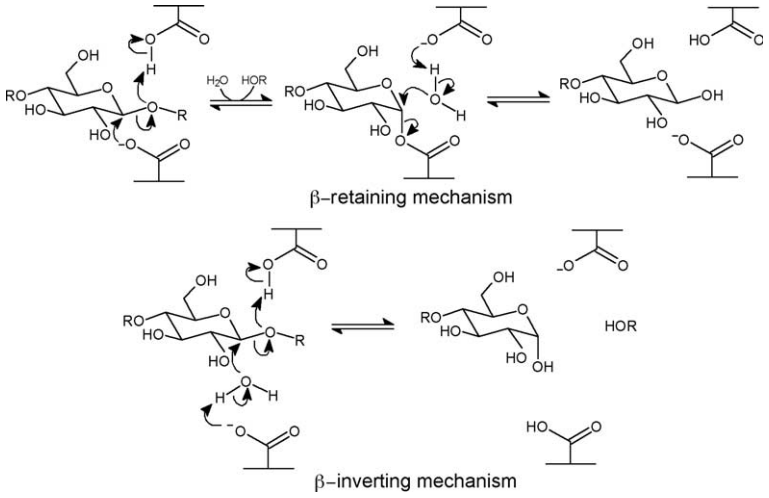

Fig. 1. The general mechanism of a β-glucoside hydrolysis by retaining and inverting enzymes, not taking into account the conformational changes during the processes.

(1) Members of this class of enzymes should be sensitive targets for transition state analogous inhibitors [13]. Indeed, several natural and synthesised nitrogen-containing sugar-analogues are known to be excellent inhibitors for selected glycoside hydrolases [14], some of which are candidates for therapeutic agents, e.g., against tumour metastasis [15].
(2) Since electrostatic factors contribute the most to transition state stabilisation [2–4], such a low $ES^{\ddagger}$-K_{diss} implies that crucial enzyme functional groups should be strategically positioned such as to favourably interact with those sectors of the carbohydrate-substrate that transiently undergo charge-distribution changes upon reaching the transition state. These groups should be prominently present within *any* glycoside hydrolase and should be readily recognisable in complexes in which the −1 subsite is occupied by a substrate (or transition state) analogue.

One perfect charge-complementarity, which is always present in enzymatic glycoside hydrolysis mechanisms, evidently consists of the protonated glycosidic oxygen of the scissile bond versus the conjugated base of the proton donating aspartic or glutamic acid residue. However, a principal characteristic of a carbohydrate ring in its transition state conformation – substantially different from that of a ground state chair or local minimum skew-boat – is the sp^2-hybridised geometry of both C1 and O5, allowing a mutual double bond character in a borderline SN_1–SN_2 displacement mechanism (Fig. 2, left). The partial positive charge at the anomeric centre is then stabilised by the ring oxygen, through overlap of O5's orthogonal lone pair containing $2p_z$ orbital with C1's orthogonal orbital lobes of the bonds from the partially leaving and partially incoming groups (or in extremis, in case of a genuine SN_1: overlap with C1's empty $2p_z$ orbital). This gives rise to two localised areas that are sterically accessible for complimentary interaction with the enzyme, i.e., axially above and underneath the carbohydrate's ring oxygen, where *at the transition state the local charge distribution differs substantially from that of the ground state*. Calculations of partial charge densities indicate

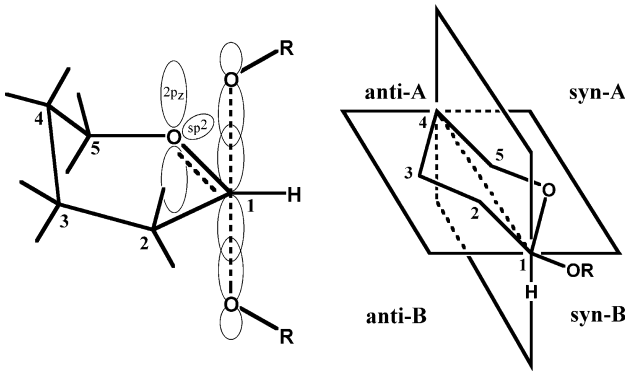

Fig. 2. Left: Schematic representation of the transition state for a glycosidic bond substitution. OR = leaving or incoming group. The ring oxygen's $2p_z$ orbital overlaps with the C1–O1 orbitals of the bonds being formed or broken. The local positive charge build-up at C1 is stabilised through this orbital overlap. Right: The *syn*-A, *syn*-B, *anti*-A, and *anti*-B space-quadrants around a glycosidic ring.

that even a formal glucopyranosyl cation may still carry a small negative charge on the ring oxygen, e.g., −0.0086, compared to −0.2607 at the ground state α-D-glucose [16]. However, the ring oxygen of a glycopyranosyl cation bears a fully occupied sp^2-hybrid and an electron-deficient $2p_z$ orbital, whereas that of a ground state (or a local minimum) bears two fully occupied sp^3 lone pairs; the ring oxygen's $2p_z$ orbital at a transition state may thus carry a substantial positive charge. A local complementary interaction from the enzyme may then be transition state stabilising as well as ground state destabilising, and as such lower the activation barrier.

A combination of Vasella's *syn–anti* half-spaces (divided by the O1, C1, and H1 plane with respect to the carbohydrate, respectively, containing O5 or C2) [9] with Vyas's A–B half-space nomenclature (with respect to the average sugar ring plane and the ring atom numbering, respectively, above-clockwise or below-anticlockwise) [17] yields the space-quadrants *syn*-A, *syn*-B, *anti*-A, and *anti*-B (Fig. 2, right). Thus, in the known 3D-structures of glycoside hydrolase enzyme–ligand

Table 1
List of 42 family/clan members of subsite −1 liganded glycoside hydrolases and their mechanistic actors

Family	Mechanism	Proton donor[a]	Nucl.[a]	*Syn*-A (Å)[i]	*Syn*-B (Å)[i]	Organism	Enzyme	PDB[n]	Ligand	Conservation of *syn*-helper residue[s]
1/A	β ret.	*anti* Glu166$^{(e)}$	Glu352$^{(e)}$	Solvent	**Tyr296** (2.8)	*B. polymixa*	1,4-Glucosidase	1e4i	GlcF-glycenz.	>95% Y, some F, G
2/A	β ret.	*anti* Glu461$^{(e)}$	Glu537$^{(e)}$	Wat4568 (2.9)	**Tyr503** (3.1)	*E. coli*	Galactosidase	1jz0	GalF-glycenz.	±60% Y, 20% E, others F, V, N
3	β ret.	*anti* Glu491$^{(e)}$	Asp285$^{(e)}$	Solvent	**Met316**[j] (4.5)	*H. vulgare*	1,4-Glucohydrolase	1iew	GlcF-glycenz.	Strict
5/A	β ret.	*anti* Glu139$^{(e)}$	Glu228$^{(e)}$	Solvent	**Tyr202** (3.0)	*B. agaradhaerens*	1,4-Endoglucanase	1h2j	GlcF-glycenz.	Strict
6	β inv.	*syn*-A Asp226$^{(e)}$	H_2O	Protond. (4.1)	Wat2902 (2.8)	*H. insolens*	1,4-Exoglucanase	1ocn	Glc-isofagomine	
7/B	β ret.	*syn*-A Glu202$^{(e)}$	Glu197$^{(e)}$	Protond. (2.9)	**Asp199** (3.1)	*F. oxysporum*	1,4-Endoglucanase	1ovw	Thio-Glc_5[p]	Strict
8/M	β inv.	*anti* Glu95$^{(e)}$[b]	H_2O	C6–OH (2.8)	Solvent	*C. thermocellum*	1,4-Exoglucanase	1kwf	Cellohexaose	
9	β inv.	*syn*-A Glu424$^{(e)}$	H_2O	Protond. (3.1)	**Asp55** (4.3)	*T. fusca*	1,4-Endoglucanase	4tf4	Cellopentaose	>95% D, some H
10/A	β ret.	*anti* Glu127$^{(e)}$	Glu233$^{(e)}$	Solvent	**His205** (3.0)	*C. fimi*	Xylanase	2xyl[o]	Xyl_2F-glycenz.	Strict (once Y)
11/C	β ret.	*syn*-A Glu172$^{(e)}$	Glu78$^{(e)}$	Protond. (4.0)	**Tyr69** (3.0)	*B. circulans*	Xylanase	1bvv	Xyl_2F-glycenz.	Strict
12/C	β ret.	*syn*-A Glu203$^{(e)}$	Glu120$^{(e)}$	Protond. (3.5)	**Asp104** (3.9)	*S. lividans*	1,4-Glucanase	2nlr	Glc_3F-glycenz.	±90% D, 10% E
13/H	α ret.	*anti* Glu257$^{(e)}$[c]	Asp229$^{(e)}$[c]	Nucl. (3.7)	Solvent	*B. circulans*	Cyclodextrin transf.	1cxk	Maltononaose	
14	α inv.	*syn*-B Glu172$^{(i)}$	H_2O	**Glu367** (3.8)	Protond. (4.5)	*B. cereus*	1,4-Glucoamylase	1b9z	Maltose	±90% E, 6% Q, 3% R
15/L	α inv.	*syn*-B Glu179$^{(e)}$	H_2O	**Tyr48** (3.6)	Protond. (3.6)	*A. awamori*	1,4-Glucoamylase	1gah	Acarbose	Strict
16/B	β ret.	*syn*-A Glu152$^{(e)}$	Glu147$^{(e)}$[h]	Protond. (3.4)	**Asp149** (3.4)	*Z. galactanivorans*	β-Agarase	1urx	Oligo-agarose	Strict
18/K	β ret.	*anti* Glu315$^{(i)}$	Internal	**Asp391** (4.4)	Nucl. (1.8)	*S. marcescens*	Chitinase	1ffr	$(NAG)_6$	>95% D, some Y, N, E, S
20/K	β ret.	*anti* Glu540$^{(i)}$	Internal	C3′-OH (3.5)	**Tyr669** (4.6)	*S. marcescens*	Chitobiase	1c7s	Chitobiose	>95% Y, some F, D, L
22	β ret.	*syn*-A Glu35$^{(e)}$[d]	Asp52$^{(e)}$	Protond. (4.3)	**Gln57**[l] (3.8)	*G. gallus*	Lysozyme	1h6m	$(Chit)_2$F-glycenz.	>95% Q, some E
23	β inv.	*syn*-A Glu73$^{(i)}$	H_2O	Protond. (3.8)	**Gln95**[l] (2.9)	*C. atratus*	Lysozyme	1lsp	Bulgecin A	±95% Q, 5% M
24/I	β inv.	*syn*-A Glu11$^{(i)}$	H_2O	Protond. (3.8)	Solvent	Bacteriophage T4	Lysozyme	148l	Glycenz.[q]	
26/A	β ret.	*anti* Glu212$^{(e)}$[e]	Glu320$^{(e)}$	Solvent	**Tyr285** (3.4)	*C. japonicus*	Mannanase	1gw1	$(Man)_3$F-glycenz.	Strict
27/D	α ret.	*anti* Asp201$^{(i)}$	Asp410$^{(i)}$	Nucl. (2.7)	Wat182 (3.6)	*G. gallus*	galactosaminidase	1ktc	NAGal	
28/N	α inv.	*anti* Asp180$^{(e)}$	H_2O	Solvent	**Tyr270** (3.2)	*A. aculeatus*	polygalacturonase	model	$(Gal)_8$	>95% Y, some W, F
29	α ret.	*syn*-A Glu266$^{(e)}$	Asp224$^{(e)}$	Protond. (3.0)	Nucl. (3.0)	*T. maritima*	Fucosidase	1hl9	FucF-glycenz.	
31	α ret.	*anti* Asp482$^{(e)}$	Asp416$^{(e)}$	**Trp345** (4.4)[j]	Solvent	*E. coli*	Xylosidase	1xsk	5FXyl-glycenz.	±49% I, 18% W, 8% N, 7% S, 6% Y, 5% M
33/E	α ret.	*anti* Asp59$^{(i)}$	Tyr342$^{(e)}$	**Glu230** (4.3)	Solvent	*T. cruzi*	Sialidase	1s0k	2,3F-sialglycenz.	Strict
34/E	α ret.	*anti* Asp151$^{(i)}$	Tyr406$^{(i)}$	**Glu277** (3.8)	Solvent	Influenza virus	Sialidase	2bat	Sialic acid	Strict
38	α ret.	*anti* Asp341$^{(e)}$	Asp204$^{(e)}$	Nucl. (2.9)	Solvent	*D. melanogaster*	1,2-Mannosidase	1qwn	5FGulF-glycenz.	
39/A	β ret.	*anti* Glu160$^{(e)}$	Glu277$^{(e)}$	Solvent	**Tyr230** (3.1)	*T. saccharolyticum*	Xylosidase	1uhv	XylF-glycenz.	±70% Y, 30% K
42/A	β ret.	*anti* Glu141$^{(i)}$	Glu312$^{(i)}$	Solvent	**Tyr266** (4.3)	*T. thermophilus*	Galactosidase	1kwk	Galactose	±90% Y, 10% F
43/F	α inv.	*anti* Glu221$^{(i)}$	H_2O	Solvent	**Cys242** (4.5)[k]	*C. cellulosa*	Arabinanase	1gye	Arabinohexaose	±24% E, 9% Y, 20% A, 10% C, 7% G
47	α inv.	*syn*-B Glu330$^{(i)}$	H_2O	C6–OH (3.5)	Wat8 (3.1)[m]	*H. sapiens*	1,2-Mannosidase	1fo3	Kifunensine	
51/A	α ret.	*anti* Glu 175$^{(e)}$[f]	H_2O	Solvent	**Tyr246** (3.1)	*G. stearotherm.*	Arabinofuranosidase	1pz2	Glycenz.	Strict
54	α ret.	*anti* Asp 297$^{(e)}$	Glu221$^{(e)}$	**Cys176** (4.4)[k]	Solvent	*A. kawachii*	Arabinofuranosidase	1wd4	L-Arabinose	Strict
56	β ret.	*anti* Glu113$^{(i)}$	Internal	**Tyr227** (5.0)	Solvent	*A. mellifera*	Hyaluronidase	1fcv	$(Hyaluron.)_4$	>95% Y, some G, C, S
57	α ret.	*anti* Asp214$^{(i)}$	Glu123$^{(i)}$	Nucl. (3.5)	Solvent	*T. litoralis*	Glucanotransferase	1k1y	Acarbose (−3/+1)	
67	α inv.	*syn*-B Glu292$^{(i)}$	H_2O	Solvent	Protond. (5.1)	*P. cellulosa*	Glucuronidase	1gql	Glucuronic acid	
68/J	β ret.	*anti* Glu342$^{(e)}$[g]	Asp86$^{(e)}$	Solvent	Nucl. (3.2)	*B. subtilis*	Levansucrase	1pt2	Sucrose	
77/H	α ret.	*anti* Asp395$^{(i)}$	Asp293$^{(i)}$	Nucl. (5.8)	Solvent	*T. aquaticus*	Amylomaltase	1esw	Acarbose (−3/+1)	
83/E	α ret.	*anti* Arg174$^{(i)}$	Tyr526$^{(i)}$	**Glu401** (4.8)	Solvent	Paramyxovirus	Sialidase	1e8v	DD-neuram.[r]	Strict

			H_2O	Solvent	Solvent	Tailspike protein	Endorhamnosidase	1tyw	Epitope fragment
90	α inv.	*anti* Asp392[(i)]	Phosphate	Protond. (3.8)	Nucl.[n]	*V. proteolyticus*	Chit. phosphorylase	1v7x	GlcNAc
94	β inv.	*syn*-A Asp492[(e)]							

[a(e)]Proton donor and/or nucleophile experimentally confirmed; [(i)]*syn* or *anti* proton donor, and nucleophile, inferred from the relative position versus the ligand.
[b]Proton donor E95Q mutant.
[c]Proton donor E257Q/nucleophile D229N double mutant.
[d]Proton donor E35Q mutant.
[e]Proton donor E212A mutant.
[f]Proton donor E175Q mutant.
[g]Proton donor E342A mutant.
[h]Nucleophile E147S mutant.
[i]Distance towards the ligand's ring heteroatom.
[j]Distance to Nε1 of Trp 345.
[k]Distance to the residue's sulfur atom.
[l]Distance to the residue's backbone carbonyl oxygen.
[m]Wat8 is held at the *syn*-B position by the proton donor E330.
[n]This structure contains a sulfate molecule instead of its phosphate nucleophile.
[o]Reference in pdb-entry 2xyl incorrect, see [31]; all other references as in the respective pdb-entry.
[p]Methyl *S*-β-D-glucopyranosyl-(1→4)-*S*-4-thio-β-D-glucopyranosyl-(1→4)-*S*-4-thio-β-D-glucopyranosyl-(1→4)-4-thio-α-D-glucopyranoside.
[q]Although an inverting enzyme, this T26E/C54T/C97A mutant forms a glycosyl-enzyme intermediate.
[r]2,3-Dehydro-2-deoxy-*N*-acetyl neuraminic acid.
[s]The helper residues at *syn*-A or *syn*-B are indicated in bold.

complexes, one expects to find at the −1 subsite and at *syn*-A and/or *syn*-B-axially above and/or underneath the ligand's ring heteroatom – always an electron-rich functional group (e.g., bearing a free electron pair), at a close enough distance to be able to interact with the substrate-ring oxygen's $2p_z$ orbital at the transition state.

This rationale has prompted us to investigate in silico, for all the to this date known liganded glycoside hydrolase 3D-structures, whether such a strategically positioned group is indeed provided (directly or indirectly) by the enzyme at the indicated *syn* quadrants. This has already been discerned and its importance deduced in liganded structures from families 11 [18], 13 [19], 29 [20] and 38 [21]; each is discussed further. The selection criteria for this study are those pdb-entries that are ligand–enzyme complexes with subsite −1 occupied, or are trapped *O*-glycosyl intermediates; when more than one structure with a ligand has been published for the same enzyme, the entry with the covalent glycosyl-enzyme intermediate (if available) was preferred. The modelled substrate complex of the GH-family 28 polygalacturonase from *Aspergillus aculeatus* [22] has also been included. No representative for GH-family 4 was included, since these enzymes operate by an unusual redox and elimination mechanism [23].

The results of the present survey for representatives from 42 different GH-families are presented in Table 1. At the *syn*-A and *syn*-B columns, the indicated distances are for the nearest axial heteroatom(s) towards the respective ligand's ring heteroatom. It must be emphasised that an experimentally observed ligand conformation must be a ground state or a local minimum, thus energetically far away from the short-lived high-energy transition state; the distances in Table 1 can therefore substantially differ from those occurring at the real transition state.

As is immediately clear from the list, only the GH-family 90 tailspike protein structure [24] is a notable exception in that its ligand's *syn*-A and *syn*-B quadrants both appear to be solvent accessible. It could be an "exception that confirms the rule", since this enzyme hydrolyses its *S. typhimurium O*-antigen substrate exceptionally slowly ($k_{cat} \pm 2$ min^{-1}) [25] and as such may not need additional electronic transition state stabilisation apart from the protonated glycosidic oxygen – deprotonated proton donor combination. A persistent occurrence of an electron-rich functional group at *syn*-A or *syn*-B also helps to explain why these enzymes consistently show a reluctance to accept a ground state substrate-carbohydrate moiety within their −1 subsite: not only is the shape of this subsite not complimentary to the ground state, but also an electrostatic mismatch is presented near the lone pairs of the ground state's ring oxygen.

It should be noted that Vasella's nomenclature for *syn–anti* proton donor positioning [9] originally was described only for β (equatorial) glycoside hydrolases. An extra result of this 3D-survey is that this appears to be equally applicable to enzymes hydrolysing an α (axial) glycosidic bond. The IUPAC β–α carbohydrate nomenclature very often (with some exceptions, see further) corresponds to an equatorial, respectively, axial, positioning of the glycosidic bond in the ground state conformation.

By nature of their positioning with respect to the substrate, *syn* versus *anti* protonators show a substantially different strategy for electronic transition state stabilisation, thus necessitating separate general discussions.

3. *Syn* protonation as a lateral phenomenon: a refinement

When a *syn* protonating glycoside hydrolase has transferred its catalytic proton to the substrate's glycosidic bond and the reaction is reaching the transition state, the acid catalyst in its basic form is then inherently also placed in the vicinity of the $2p_z$ orbital of the substrate's sp^2-hybridised endocyclic oxygen. A transition state favouring interaction towards this orbital should only be possible with *syn* proton donor positions that are axial to the ring oxygen of the carbohydrate-unit at the −1 subsite. These are exactly the *syn*-A and *syn*-B quadrants as indicated above (Fig. 2, right). If, on the contrary, the proton donor would actually reside in the sugar's average ring plane, the interaction would even be disadvantageous since its conjugate base would be very close to the ring oxygen's fully occupied equatorial lone-pair sp^2-hybrid orbital. This can be regarded as a refinement of Heightman and Vasella's original insight of lateral proton donor positioning [9]: it is semi-lateral, with the carbonylic oxygen of the proton donor axially positioned versus the ring oxygen and with the protonated oxygen of the proton donor in the vicinity of the glycosidic oxygen.

The present survey indeed reveals (see the syn proton donor entries in Table 1) that *syn*-β (equatorial) glycoside hydrolases consistently have their proton donor at the *syn*-A quadrant, whereas *syn*-α (axial) enzymes show this in the corresponding *syn*-B quadrant. The latter is logical, since here the axial glycosidic oxygen that is to be protonated resides in the B half-space, but the consistent presence of a proton donor in the *syn*-A quadrant with *syn* protonating β-glycoside hydrolases leads to a further interesting possibility. Indeed, a ground state chair conformation of an equatorial glycoside could in principle be protonated semi-laterally, followed by conjugate base interaction to the ring oxygen from either *syn*-A or *syn*-B. However, the antiperiplanar lone pair hypothesis (ALPH) predicts that the hydrolysis of equatorial glycosides should be preceded by a conformational change away from the ground state chair to a twist-boat conformation in which the leaving group has an axial position, enabling a set-up with an antiperiplanar orbital of the ring oxygen lone pair – the one that will eventually become the $2p_z$ orbital at the transition state [26–28]. The observation that the proton donor of *syn* protonating (equatorial) β-glycoside hydrolases is always positioned at *syn*-A suggests that their protonation of the glycosidic oxygen does not occur at the ground state chair conformation, but always happens after a conformational change to an axial anomeric conformer. This may or may not be so for *anti* protonating β-glycoside hydrolases, since their proton donor appears to occupy a more variable position.

In general, since *syn* protonators have their proton donor positioned in the *syn*-A or *syn*-B quadrant and as such are pre-set for an extra interaction towards the transition state's ring oxygen, they should be in catalytic advantage over *anti* protonators. This fits the interesting observation that inverting β-glycoside hydrolases, using water as a nucleophile (which is less nucleophilic than a carboxylate-anion used in the rate-determining step of retaining enzymes), are always *syn* protonators with the exception of the members of clan M (families 8 and 48). This *anti* protonating GH-clan M is again peculiar in that the known 3D-structures show a very extended multi-subsite active centre, as described further.

4. *Anti* protonators should compensate at a *syn* quadrant

Since the conjugate base of an *anti* positioned proton donor cannot interact with the ring oxygen's $2p_z$ orbital of the transition state, one expects *anti* protonating glycoside hydrolases to provide an alternative interaction originating from the *syn*-A and/or *syn*-B quadrants. As apparent in Table 1 (see the *anti* proton donor entries), an electron-rich heteroatom is always present (except in family 90). In two instances (families 8 and 47), a hydroxyl group from the ligand itself is located at the correct distance; such interaction is then indirectly provided by the enzyme's active site, by specifically locking its substrate into that position. Five α-retaining families (13, 27, 38, 57, and 77) show the second oxygen of their carboxylate-nucleophile at *syn*-A. Most *anti* protonating enzymes provide this axial interaction by a separate amino acid residue (boldfaced in Table 1).

Many *syn*-protonating glycoside hydrolases also show a second electron-rich amino acid residue (boldfaced in Table 1) located at the *syn*-quadrant opposite to that containing the proton donor. A check with the PFAM sequence-alignment server [29] or with ClustalW [30] shows that the residues involved, for *anti* as well as *syn* protonating enzymes, are often highly conserved within each family. This indicates that these may be regarded as important transition state favouring helper residues. As the examples show, known mutations of this residue are often deleterious for enzyme activity; for some mutants however the effects are less pronounced.

5. Specific examples for different mechanistic combinations

A description of all entries in Table 1 is impossible within the scope of the present article. The permutation of (i) axially versus equatorially positioned glycosidic bond substrates, (ii) retaining versus inverting hydrolases, and (iii) *syn* versus *anti* proton donor positions yields eight mechanistic combinations; for a reasonable completeness a selection of examples is given for each case. The GH-clan E families 33, 34, and 83 sialidases-neuraminidases are discussed in a separate section, since their substrates possess an equatorial anomeric substituent although the IUPAC nomenclature requires the α denominator.

The present focus is only on enzyme-provided electrostatic factors within the *syn*-A and *syn*-B quadrants. Undoubtedly, many other enzyme-provided features will also contribute to transition state stabilisation, e.g., specific nucleophile to the C2-hydroxyl group interactions often found in retaining enzymes [31,32], or hydrophobic platform interactions [33].

The wall-eyed stereoviews of Fig. 3 (*syn*-β-retaining enzymes) and Fig. 4 (*anti*-β-retaining enzymes) are pdb-extracts of the liganded −1 subsite from several glycoside hydrolase family members, showing an overlay of the proton donors and the amino acid residues under the present focus. The manual overlays, by coalescence of the respective C1–O5 bonds and subsequent best fit of the carbohydrate rings at the −1 subsite, were performed with the Swiss-PDB-viewer 3.7 [34] and the final figures were prepared with Pymol-X11-hybrid 0.95 [35].

5.1. Syn β-retaining: GH-clans B and C (Fig. 3)

The clan B β-glycoside hydrolases operate with retention of the equatorial anomeric configuration and exhibit a common

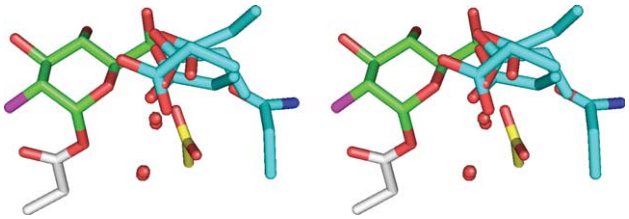

Fig. 3. Wall-eyed stereoview overlay of the situations at the liganded active sites from the *syn*-β-retaining families 7 (1ovw), 11 (1bvv), 12 (2nlr), 16 (1urx), and 22 (1h6m). Only the 2-fluoro-glycosyl-enzyme intermediate from family 12 is shown (green, fluorine atom in purple), attached to the nucleophile Glu120 (white), and its helper residue Asp104 (yellow) in the *syn*-B quadrant. The respective proton donors (blue) all reside in the *syn*-A quadrant; the proximal oxygen atoms (red) of the respective *syn*-B helper residues are represented as coloured spheres.

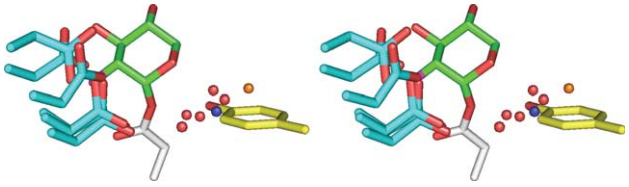

Fig. 4. Wall-eyed stereoview overlay of the situation at the active sites of glycosyl-enzyme intermediates from the *anti*-β-retaining GH-clan A families 1 (1e4), 2 (1jz0), 5 (1h2j), 10 (2xyl), 26 (1gwi), 39 (1uhv), and the structurally analogous (GH-clan A) *anti*-α-retaining L-arabino-furanosidase from family 51 (1pz2); the glycosyl-enzyme intermediate from family 3 (1iew) was also included. Only the 2-fluoro-glycosyl-enzyme intermediate from family 39 is shown (green, fluorine atom in purple), attached to the nucleophile Glu277 (white), and its helper residue Tyr230 (yellow) in the *syn*-B quadrant. The respective proton donors (blue) are *anti* positioned; the proximal heteroatoms (red, oxygen; blue, nitrogen; and orange, sulfur) of the respective *syn*-B helper residues are represented as coloured spheres.

β-jelly-roll fold with an extended multi-subsite active centre. Cellulose (family 7) or complex polysaccharides (family 16) are their substrates. 3D-Complexes in which subsite −1 is occupied (several pdb-entries for family 7, only entry 1urx for family 16) show the glutamic acid proton donor in the *syn*-A quadrant, whereas in both families at *syn*-B a strictly conserved aspartate residue is present which is also close to the glutamate nucleophile. The selected example in Table 1 is the *Fusarium oxysporum* endoglucanase Cel7B in complex with a non-hydrolysable thiopentasaccharide substrate-analogue inhibitor (pdb-entry 1ovw) [36]. The oxygen Oε2 of the proton donor Glu202 is at *syn*-A with a distance of 2.9 Å to the ring oxygen of the substrate analogue in the −1 subsite, and oxygen Oδ1 of Asp199 resides clearly at *syn*-B at 3.1 Å distance. With the *Hypocrea jecorina* cellobiohydrolase Cel7A, the amide mutant of the corresponding *syn*-B positioned Asp214 shows on aryl glycoside substrates a two orders decrease in k_{cat} and an unchanged K_M. Its possible role as a nucleophile assisting residue – but not its proximity to the ring oxygen – has been extensively discussed [37].

The β-retaining GH-clan C also contains β-jelly-roll folded enzymes with an extended multi-subsite active centre, and comprises the family 11 xylanases and the family 12 endoglucanases and lichenases.

The selected example in Table 1 for family 12 is the covalent 2-deoxy-2-fluoro-cellotriosyl-enzyme intermediate of the CelB2 endoglucanase from *Streptomyces lividans* (pdb-entry 2nlr) [38]. At *syn*-A, one finds the oxygen Oε2 of the proton donor Glu203 at 3.5 Å distance to the ligand's ring oxygen, whereas in the *syn*-B quadrant at 3.9 Å the carboxyl oxygen Oε1 of Asp104 is observed. This residue is 85% conserved with 15% Glu as variant; mutants at this position have not yet been reported.

The GH-family 11 *syn* protonating xylanases show the phenolic oxygen of a strictly conserved tyrosine residue at *syn*-B, in a near axial position versus their substrate's ring oxygen; a feature that is shared by many other families. The selected example in Table 1 is the covalent 2-deoxy-2-fluoro-xylobiosylenzyme intermediate of the *Bacillus circulans* xylanase (pdb-entry 1bvv) [18]. Remarkable in this structure, and also observed in a covalent intermediate of the family 11 *Bacillus agaradhaerens* xylanase [39], is the presence of a ^{2,5}B boat conformation of an α-D-xyloside moiety in the −1 subsite. This conformation is coplanar with respect to C5, O5, C1, and C2, conform to what one expects from a transition state that has considerable double bond character between O5 and C1. Apparently in contrast to ALPH, the ring oxygen provides no lone pair that is antiperiplanar to the α-axial glycosidic bond. However, a minimal disturbance of this ^{2,5}B boat conformation yields a $^{5}S_{1}$ skew-boat as the next pseudorotational neighbour which now is ALPH compliant and could be the actual active conformation. Also, the −1 subsite of these enzymes appears to possess an unusual preference for a local minimum α-axial ^{2,5}B boat D-xylopyranose conformation, since this is also observed in the xylotetraose product-complex of the *B. agaradhaerens* xylanase nucleophile Glu94Ala mutant [39]. In the *B. circulans* xylanase, one finds at *syn*-B the phenolic oxygen of Tyr69 at 3.0 Å from the ring oxygen; the Tyr69Phe mutant yields an inactive enzyme [40]. Its importance in transition state stabilisation is thoroughly discussed by the authors of the 3D-structure [18] and is in accordance with the general proposition of the present review.

An identically *syn*-B positioned tyrosine helper residue is also found within the β-retaining but *anti* proton donating GH-family 20 and in most families of the GH-clan A (see further). An extra feature of a *syn*-B tyrosine with a β-retaining enzyme combination is that the tyrosine's phenolic group inherently is also positioned in the vicinity of the nucleophile. As such, it can yield a bifurcated hydrogen bond towards the nucleophile as well as towards the endocyclic ring-oxygen of the substrate. The changes in charge distribution in going from a ground state towards the transition state could favour the hydrogen bond towards the nucleophile, while exposing the substrate's ring oxygen more prominently to a lone pair from the phenolic oxygen, which in turn could stabilise the transition state. This has indeed been observed and discussed for the family 11 *B. circulans* xylanase [18] and may be a recurring phenomenon with many other glycoside hydrolases.

5.2. Anti β-retaining: GH-clan A (Fig. 4)

The large GH-clan A comprises to this date 17 different families, all sharing a common $(\beta/\alpha)_8$-fold and acting by a retaining mechanism; the complexed structures all show an *anti* positioned proton donor. GH-family 5 enzymes hydrolyse various β-(1,4)-glycosides with an equatorial anomeric configuration. The selected example in Table 1 is the 2-deoxy-2-fluoro-cellobiosylenzyme intermediate of the endoglucanase

Cel5A from *B. agaradhaerens* (pdb-entry 1h2j) [41]. In the *syn*-B quadrant, at 3.0 Å distance from the ligand's ring oxygen, one finds the phenolic oxygen of Tyr202; the *syn*-A quadrant is solvent exposed. The substrate specificities of GH-family 5 enzymes are broad and some members share a mutually closer sequence identity, which has led to a subdivision into as yet eight subfamilies [42]. Nevertheless, the indicated tyrosine is strictly conserved.

As indicated in Table 1, a tyrosine at this position is found throughout the 3D-structures of all families in the GH-clan A, except in the family 10 xylanases and cellobiohydrolases where at the *syn*-B quadrant the nitrogen atom of a conserved histidine is present, although one enzyme sequence again shows a tyrosine. As required for transition state stabilisation, the imidazole nitrogen atom near the substrate's ring oxygen (e.g., Nε2 of His205 in the *Cellulomonas fimi* example) should be basic, and indeed the other nitrogen (e.g., Nδ1) carries the proton since here an adjacent aspartate residue (e.g., Asp235) is invariantly present [31]. In the GH-clan A family 39, the indicated tyrosine is 70% conserved next to 30% lysine; the latter is then predicted to also have a neighbouring carboxylate residue. Within the GH-clan A, the indicated tyrosine is >95% conserved in families 1, 5, 10, 26, and 51, whereas this is 90% in family 42 (10% Phe), and 60% in family 2 (20% Glu and some Phe and Gly). The occurrence of a phenylalanine at this position is remarkable and suggests that a polarisable π-aromatic ring system may also help in transition state stabilisation; a glycine at this position may be due to a sequence error or may indicate that a neighbouring residue occupies the *syn*-B quadrant.

Mutants have been characterised for several enzymes belonging to GH-clan A. The Tyr298Phe mutant of the family 1 β-glucosidase from *Agrobacterium faecalis* shows a 100- to 3000-fold decrease of k_{cat} and a virtually unchanged to 30-fold lower K_M on different aryl glucosides [43]. With the family 1 human cytosolic β-glucosidase, the Tyr308Phe mutant exhibits however only a 2–5-fold decrease of k_{cat} towards its flavanoid glucoside substrates with minimal influence on K_M, which indicates that a polarisable π-aromatic ring system may have a capacity for transition state stabilisation; it also may be a reflection of this enzyme's activity on a variety of phenolic glucosides [44]. The family 2 β-galactosidase Tyr503Phe mutant from *E. coli* shows a 2800, respectively, 900-fold decrease of V_{max} for *o*-nitro-, respectively, *p*-nitrophenyl β-D-galactoside, with a 3-fold decrease of K_M [45]. The Tyr285Ala mutant of the family 26 mannanase from *Pseudomonas cellulosa* exhibits a four orders of magnitude decrease of k_{cat} and a 4-fold increase of K_M on its natural substrate [46].

5.3. Syn β-inverting: family 9

This family contains endoglucanases, cellobiohydrolases and β-D-glucosidases, acting on substrates with an equatorial anomeric configuration. In the *Thermomonospora fusca* Cel9A cellopentaose complex (Table 1, pdb-entry 1tf4) [47], the proton donor Glu242 is clearly occupying the *syn*-A quadrant, and its carboxylic oxygen Oε1 is at an appropriate distance (3.1 Å) from the ligand's ring oxygen for transition state electrostatic interaction by its conjugate base. At the B half-space near the ligand's anomeric centre, two adjacent and probably cooperative bases Asp58 and Asp55 are present. The carboxylic oxygen Oδ2 of the latter is positioned axially at 4.3 Å distance from the carbohydrate's ring oxygen. This residue is strictly conserved, and mutants (D55N as well as D55A) show a 100-fold decrease in k_{cat} whereas K_M is almost unaffected [48].

5.4. Anti β-inverting: family 8 (clan M)

As indicated above, *anti* proton donating glycoside hydrolases are in relative disadvantage versus *syn* protonating enzymes, and are expected to compensate for the absence of the proton donor's conjugate base proximity towards the ring oxygen of the carbohydrate-unit at the −1 subsite by locally providing an alternative transition state favouring inter-actor. Equatorial glycosidic bond hydrolysing enzymes may have a second disadvantage due to ALPH, which suggests the need for a prior substrate conformational change (see above). The required energy for converting a ground-state chair conformation into a local minimum ALPH-compliant (skew-)boat must come from an extra series of favourable enzyme–substrate interactions. And, since inverting enzymes utilise water as a nucleophile, *anti*-β-inverters accumulate a 3-fold-cumulative disadvantage, and should be very rare in Nature. Surprisingly, the GH-families 8 and 48 enzymes are *anti*-β-inverting. An example is the active site of the *Clostridium thermocellum* endoglucanase Cel8A (Table 1, pdb-entry 1kwf, substrate complex with proton donor E95Q mutant) [49]. It shows the D-glucose moiety in subsite −1 residing in a non-ground state $^{2}S_{0}$-^{2,5}B conformation. This is a non-ALPH compliant local minimum conformation which could be due to the mutation, whereas in the native enzyme the substrate might first pseudorotate to an ALPH-compliant conformation before the actual hydrolysis reaction. The proton donor is clearly *anti* positioned and at *syn*-A one finds the D-glucose's C6 hydroxyl oxygen at 2.8 Å distance from the ring oxygen. Its position appears to be enforced by a steric interaction from Trp132, which also functions as a subsite −2 ligand stacking residue.

The clan GH-clan M families 8 and 48 are $(\alpha/\alpha)_6$ barrel folded and show a very extended active site. Family 8 glycosidases are only active on oligosaccharides with a high degree of polymerisation [49,50], whereas 3D-structures from members of the processively hydrolysing family 48 show a long and elaborate tunnel-shaped active site (e.g., the *Clostridium cellulolyticum* Cel48A structure [51]). Utilising many subsites with favourable enzyme–substrate interactions may be Nature's solution for these enzymes' inherent mechanistic disadvantages.

5.5. Syn α-retaining: family 29

For this family of α-L-fucosidases (its substrates having an axial anomeric configuration), only 3D-structures of the *Thermotoga maritima* enzyme have as yet been solved, including that of its 2-deoxy-2-fluoro-L-fucosylenzyme intermediate (Table 1, pdb-entry 1hl9) [20]. The Oε2 carbonyl oxygen of its proton donor Glu266 clearly resides in the *syn*-A quadrant, at 3.1 Å distance from the ligand's ring oxygen. At *syn*-B, the second oxygen Oδ2 of the nucleophile Asp224 is found at 3.0 Å distance. The authors have identified this positioning as "possibly providing electrostatic stabilisation at the transition state, or ground state destabilisation" [20], in accordance with our general hypothesis. As evident from Table 1, a nucleophile positioned at *syn*-B also occurs in the β-inverting family 24 and in the β-retaining family 68.

The same proximity of the nucleophile's carbonyl oxygen to the ligand's ring oxygen has previously been noted with *anti* protonating α-retaining glycosidases, described in the next section.

5.6. Anti α-retaining: families 13, 27, 31, 38, 57, and 77

As noted above, these *anti* protonating enzymes cannot compensate the changes in charge distribution at their substrate's ring oxygen through interaction with the conjugate base of their proton donor. They are retainers for axially substituted substrates and are as such equipped with an inherently close anionic nucleophile. The second oxygen of the nucleophile then becomes a candidate for electronic transition state stabilisation, since a simple *syn*-A orientation would already provide a helping role in catalysis.

Table 1 indeed indicates that many known liganded 3D-structures of *anti*-α-retaining glycoside hydrolases (except a family 31 representative as discussed further and family 54 arabinoside hydrolases as discussed in Section 5.8) show the second oxygen of their nucleophile clearly situated at the *syn*-A quadrant. An excellent example is fully described in Nuamo et al.'s [21] structural elucidation of covalent intermediates with the family 38 Golgi α-mannosidase II from *Drosophila melanogaster* and includes an adequate interpretation of the peculiar position of the nucleophile.

Interestingly, fluoro derivatives of glycosyl-enzyme intermediates of this inverting α-mannosidase show an ALPH-compliant $^{1}S_{5}$ skew-boat conformation. ALPH-compliancy is also observed in the Michaelis complex of the retaining family 26 β-mannanase from *Pseudomonas cellulosa*, although here the covalent intermediate shows a $^{O}S_{2}$ conformation [52]. This suggests that the cleavage of β-mannosides follows a $^{1}S_{5} \rightarrow B_{2,5} \rightarrow {}^{O}S_{2}$ conformational itinerary in comparison to a $^{4}C_{1} \rightarrow {}^{4}H_{3} \rightarrow {}^{1}S_{3}$ itinerary for β-glucosides, while the corresponding α-glycosides – in agreement with the principle of microscopic reversibility – follow the respective inverted itineraries [21,52].

In the Michaelis complex of the retaining family 13 *B. circulans* cyclodextrin glycosyltransferase (Table 1, pdb-entry 1cxk, proton donor E257Q and nucleophile D229N double mutant), Gln257 is clearly *anti* positioned and the Asn229 amide nitrogen is at *syn*-A, axially at 3.7 Å distance from the glucoside ring oxygen. The authors' interpretation of the nucleophile positioning is in accordance with our general hypothesis [19]. The D-glucosyl unit in subsite −1 resides in a slightly distorted but ALPH-compliant $^{4}C_{1}$ conformation. The same study also describes the 4-deoxymaltotriosyl-enzyme intermediate (pdb-entry 1cxl, proton donor E257Q mutant), which surprisingly shows the −1 subsite again occupied by a ground state $^{4}C_{1}$ glucoside, now with a β-equatorial anomeric bond (non-ALPH compliant). It could however represent a non-productive local minimum, and the authors [19] suggest a strategic reason, i.e., that this could prolong the lifetime of the intermediate, thereby allowing time for a new acceptor to diffuse into the active site. This may be a recurrent strategy for glycosyl transferases.

The recently determined structure of the retaining family 31 α-D-xylosidase *YicI* from *E.coli* trapped as a 5-fluoroxylopyranosyl-enzyme intermediate (Table 1, pdb-entry 1xsk) [53] shows an *anti* positioned Asp482 as the proton donor. The nitrogen Nε1 of Trp345 is located in the *syn*-A quadrant, at 4.4 Å distance from the xyloside ring oxygen. The free electron pair of this nitrogen is however involved in the tryptophane's π-aromatic ring system and is therefore not available for direct interaction towards the substrate's ring oxygen, again suggesting that a nearby positioned polarisable π-aromatic system may be advantageous for transition state stabilisation. Remarkably, this tryptophane residue is only for 18% conserved, next to 49% Ile, 8% Asn, 7% Ser, 6% Tyr, and 5% Met. Especially the major occurrence of isoleucine at this position is intriguing, and in these cases a "normal *anti*-α-retaining situation", i.e., the nucleophile's carbonyl oxygen occupying the *syn*-A quadrant, could be expected.

5.7. Syn α-inverting: family 15 (clan L)

The inverting family 15 glucoamylases and glucodextranases, and the family 65 glycoside phosphorylases act on glycosides having an axial anomeric substituent. They show a similar $(\alpha/\alpha)_6$-fold and are as such classified in the GH-clan L. A representative structure is that of the glucoamylase from *Aspergillus awamori* var. X-100 in complex with acarbose (Table 1, pdb-entry 1gah), with the reducing-end valienamine occupying the −1 subsite [54]. The proton donor Glu179 clearly resides in the *syn*-B quadrant, and its carboxylate-oxygen Oε2 is axially at 3.6 Å distance from the endocyclic C6 of the valienamine moiety, i.e., the position of O5 of its natural substrate. In the *syn*-A quadrant, at 3.6 Å from C6, one finds the phenolic oxygen of the invariant Tyr48 residue.

5.8. Arabinoside hydrolases. Anti α-inverting: family 43 (clan F) and anti α-retaining: family 54

This glycoside hydrolase activity is observed within the structurally diverse families 3, 51, and 54, and in the GH-clan F families 43 and 62. No (liganded) structure has as yet been solved for family 62, but all others are clearly *anti* protonators. A surprising similarity at the −1 subsite of liganded family 43 and 54 structures demands a comparative description.

The inverting family 43 contains xylanases, β-D-xylosidases, arabinanases and α-L-arabinofuranosidases. As yet only the structure of the arabinase Arb43A from *Cellvibrio japonicus* has been solved, revealing a five-bladed β-propeller fold [55]. In the substrate complex with an inactive D158A mutant (Table 1, pdb-entry 1gye), one finds at the *syn*-B quadrant the sulfur atom of Cys242 at 4.5 Å distance from the endocyclic ring oxygen; this sulfur is involved in a cystine bridge with an adjacent Cys241 residue. Cys242 is only for 10% conserved within the known family 43 sequences. The other residues at this position mainly are Glu (24%) and Tyr (9%), which may reflect the different substrate specificities and the unknown function of most of the family 43 protein sequences, and surprisingly also Ala (20%) and Gly (7%), which may be due to sequence errors or may indicate that a neighbouring residue occupies the *syn*-B quadrant.

Remarkably, in the recently solved L-arabinose complex structure of the retaining family 54 α-L-arabinofuranosidase from *Aspergillus kawachii* (pdb-entry 1wd4) [56], an analogous disulfide bridge between Cys176–Cys177 is found, which is strictly conserved. The sulfur atom of Cys176 is now residing in the *syn*-A quadrant at 4.4 Å distance from the endocyclic ring oxygen. The adjacent Cys177 may be involved in a hydrophobic interaction with C5 of the ligand, and the double alanine mutant shows an increased K_M (>10 versus wt 0.76 mM) and decreased k_{cat} (3.9 versus wt 26.8 s^{-1}) for *p*-nitrophenyl

α-L-arabinofuranoside [56]. The catalytic domain of family 54 is completely different from that of family 43 and consists of a β-sandwich fold.

5.9. Families 33, 34, and 83 (clan E)

The retaining clan-E exo-α-sialidases hydrolyse sialic acid from glycoproteins. Although the scissile glycosidic bond in these substrates is positioned equatorially for the sialic acid's $^2C_5$ ground state conformation, the trans positioning within a Fisher projection of the C2 anomeric substituent versus the anomeric reference atom C7 requires this configurational relationship to be named α.

Liganded 3D-structures all show at the −1 subsite an invariant tyrosine-glutamate combination, where the tyrosine's phenolic oxygen – and not the glutamate's carboxylate – is correctly positioned for nucleophilic displacement towards the substrate. A crystal structure of the GH-family 33 *trans*-sialidase from *Trypanosoma cruzi* has been published, with this Tyr342 indeed being covalently trapped by 2,3-difluorosialic acid (pdb-entry 1s0k) [57,58]. The adjacent Glu230 clearly shows its carboxylate group situated at *syn*-A, axially versus the ligand's ring oxygen. As indicated in Table 1, the corresponding glutamates in liganded examples of sialidase families 34 and 83 also reside in the *syn*-A quadrant.

6. *Syn* protonators revisited

From the enzyme list in Table 1, one cannot fail to notice that 26 of the 42 selected GH-families are *anti* protonators, although these should be in mechanistic disadvantage in comparison to *syn* protonators which can provide an interaction of the proton donor's conjugate base to the ring oxygen's $2p_z$ orbital at the transition state. The intriguing question of why so many *anti* protonators do exist leads to a further consideration.

For a non-enzymatic glycosidic bond hydrolysis (e.g., catalysed by a dilute hydrogen chloride solution), the reaction-yielding protonation of the glycosidic oxygen is expected to occur on the lone pair orbital lobe that is involved in the exo-anomeric effect. This effect favours a O5–C1–O1–Cx dihedral angle Φ^{O5} in the range of −71° to −105° or 75° to 121°, respectively, for an equatorial or an axial glycosidic bond in glycopyranosides and oligosaccharides [59], and is mainly due to a hyperconjugative overlap of the C1–O5 antibonding orbital with an antiperiplanar oriented lone pair orbital lobe of the glycosidic oxygen, and this lobe resides in the *anti* half-space. It is a substantial stabilising effect in the order of 4 kcal/mol [60], and protonation of the involved lone pair automatically removes the stabilising effect en route to the transition state. This suggests an interesting scenario in early evolution of glycosidases. Very primordial glycosidases must have been using a slightly improved variant of a non-enzymatic acidic catalysed glycosidic bond hydrolysis mechanism. This indicates that the very first glycoside hydrolases were all *anti* protonators. Soon thereafter, evolutionary pressure for efficiency has equipped those early *anti* protonating glycosidases with a *syn* helper residue, and later some of these would start to use the *syn* helper residue as the actual proton donor because of its mechanistic advantage in transition state stabilisation.

If this scenario is correct, then at the "ex-*anti* proton donor" position of *syn* protonators one might find a remnant of the now defunct original proton donor (like the degenerate bones of hind legs in whales). In the long course of evolution however, this residue might have shifted, been given another function, or even be completely deleted. Since the least drastic switch-off of a carboxylate proton donor is by mutation to its corresponding amide, one thus expects to often find a glutamine or asparagine residue at or nearby the *anti* half-space of *syn* protonating glycoside hydrolases. This is a bold prediction coming from a speculative scenario, however, as the list in Table 2 shows, it surprisingly is the case for no less than 11 out of the 15 liganded *syn* protonating families. A typical example is the *anti* position of Gln175 in the *F. oxysporum* endoglucanase Cel7B thiooligosaccharide complex (pdb-entry 1ovw) [36]. A logical further prediction is that if one would point-mutate the indicated *anti* residue back to a carboxylate while as well mutating the *syn* proton donor to a glutamine or asparagine, one might obtain a still reasonably active but *anti* protonating enzyme. The GH-family 29 enzymes apparently lack such a remnant; here it might have been completely deleted. In the

Table 2
Liganded *syn*-protonating GH-family/clan members and their putative ex-*anti* proton donor

Family	Mechanism	Proton donor[a]	Nucleophile[a]	PDB[d]	*Anti* remnant	Conservation of *anti* remnant
6	β inv.	*syn*-A Asp226$^{(e)}$	H_2O	1ocn	Asn310	Invariant
7/B	β ret.	*syn*-A Glu202$^{(e)}$	Glu197$^{(e)}$	1ovw	Gln175	±3/4 conserved, ±1/4 Asn
9	β inv.	*syn*-A Glu424$^{(e)}$	H_2O	4tf4	His125	±1/2 Asn, ±1/5 His, ±1/5 Ser
11/C	β ret.	*syn*-A Glu172$^{(e)}$	Glu79$^{(e)}$	1bvv	Gln127	Invariant
12/C	β ret.	*syn*-A Glu203$^{(e)}$	Glu120$^{(e)}$	2nlr	Asn155	±2/3 conserved, often Gly
14	α inv.	*syn*-B Glu172$^{(i)}$	H_2O	1b9z	Asn94	Invariant
15/L	α inv.	*syn*-B Glu179$^{(i)}$	H_2O	1gah	Glu179	Asp and/or loop change may yield *anti*
16/B	β ret.	*syn*-A Glu152$^{(e)}$	Glu147$^{(e)c}$	1urx	Gln183	±1/2 conserved, others are variable
22	β ret.	*syn*-A Glu35$^{(e)b}$	Asp52$^{(e)}$	1h6m	Asn44	±3/4 conserved, often Ser
23	β inv.	*syn*-A Glu73$^{(i)}$	H_2O	1lsp	Asn148	±3/4 conserved, often Ala, His, Tyr
24/I	β inv.	*syn*-A Glu11$^{(i)}$	H_2O	148l	Gln105	±2/3 Asn, often Gln, Lys, Arg
29	α ret.	*syn*-A Glu266$^{(e)}$	Asp224$^{(e)}$	1hl9	Met225	±3/4 Gly, also Met, Trp, Phe
47	α inv.	*syn*-B Glu330$^{(i)}$	H_2O	1fo3	Glu330	Asp and/or loop change may yield *anti*
67	α inv.	*syn*-B Glu292$^{(i)}$	H_2O	1gql	Glu292	Asp and/or loop change may yield *anti*
94	β inv.	*syn*-A Asp492$^{(e)}$	Phosphate	1v7x	Gln690	±2/3 conserved, often Ala

[a](e)Proton donor and/or nucleophile experimentally confirmed, (i)*syn* proton donor inferred from the relative position versus the ligand.
[b]Proton donor E35Q mutant.
[c]Nucleophile E147S mutant.
[d]References as in the respective pdb-entry.

syn-α families 15, 47, and 67, a simple Asp to Glu mutation and/or a small change in the enzyme's local backbone might have yielded the transformation from an *anti* to a *syn* protonator.

Elaborating further on the evolutionary scenario, the first *syn* protonating glycoside hydrolases must have been faced with the problem of an again fully occupied glycosidic oxygen's lone pair that re-establishes the exo-anomeric effect. The obvious solution is an adjustment of the shape of the active site in such a way that their substrate fits in a subsite −1 to +1 carbohydrate to carbohydrate orientation, in which this lone pair is no longer antiperiplanar to the C1–O5 bond. This is indeed observed in liganded complexes of *syn* protonators where both these subsites are occupied, e.g., the *syn*-β family 7 example from Table 2 (pdb-entry 1ovw) in complex with a thiooligosaccharide shows an O5–C1–S1–C4 dihedral angle Φ^{O5} of −154°, whereas in the *syn*-α family 14 example (pdb-entry 1b9z) the orientation of the D-glucose moiety in subsite −1 versus the one occupying +1 is such that the O5–C1–O1–C4 dihedral angle Φ^{O5} would have been practically a full 180° if the interconnecting glycosidic bond had been present.

Acknowledgement: Tom Desmet holds a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

References

[1] Pauling, L. (1946) Molecular architecture and biological reactions. Chem. Eng. News 24, 1375–1376.
[2] Warshel, A. (1978) Energetics of enzyme catalysis. Proc. Natl. Acad. Sci. USA 75, 5250–5254.
[3] Warshel, A. (1981) Electrostatic basis of structure–function correlation in proteins. Acc. Chem. Res. 14, 284–290.
[4] Warshel, A. (1991) Computer Simulation of Chemical Reactions in Enzymes and Solutions, John Wiley & Sons, New York.
[5] Coutinho, P.M. and Henrissat, B. (2002) Carbohydrate-active enZYmes server. Available from: http://afmb.cnrs-mrs.fr/cazy/CAZY/index.html.
[6] Henrissat, B. and Davies, G. (1997) Sequence and structural classification of glycoside hydrolases. Curr. Opin. Struct. Biol. 7, 637–644.
[7] Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316, 695–696.
[8] Koshland, D.E. (1953) Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. Camb. Phil. Soc. 28, 416–436.
[9] Heightman, T.D. and Vasella, A.T. (1999) Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidases. Angew. Chem. Int. Ed. Engl. 38, 751–770.
[10] Zechel, D. and Withers, S.G. (1999) Mechanisms of glycosyl transfer, in: Comprehensive Natural Products Chemistry (Barton, D., Nakanishi, K. and Poulter, Eds.), vol. 5, Chapter 12, pp. 279–314, Elsevier, Amsterdam.
[11] Rye, C.S. and Withers, S.G. (2000) Glycosidase mechanisms. Curr. Opin. Chem. Biol. 4, 573–580.
[12] Vasella, A., Davies, G.J. and Böhm, M. (2002) Glycosidase mechanisms. Curr. Opin. Chem. Biol. 6, 619–629.
[13] Wolfenden, R., Lu, X. and Young, G. (1998) Spontaneous hydrolysis of glycosides. J. Am. Chem. Soc. 120, 6814–6815.
[14] Stütz, A.E. (1999) Iminosugars as Glycosidase Inhibitors. Nojirimycin and beyond, Wiley-VCH, New York.
[15] Nishimura, Y. (2003) Gem-diamine 1-N iminosugars and related iminosugars, candidate of therapeutic agents for tumor metastasis. Curr. Top. Med. Chem. 3, 575–591.
[16] Kajimoto, T., Liu, K.K.-C., Pederson, R.L., Zhong, Z., Ichikawa, Y., Porco, J.A. and Wong, C.-H. (1991) Enzyme-catalyzed aldol condensation for asymmetric synthesis of azasugars: synthesis, evaluation, and modeling of glycosidase inhibitors. J. Am. Chem. Soc. 113, 6187–6196.
[17] Vyas, N.K. (1991) Atomic features of protein–carbohydrate interactions. Curr. Opin. Struct. Biol. 1, 732–740.
[18] Sidhu, G., Withers, S.G., Nguyen, N.T., McIntosh, L.P., Ziser, L. and Brayer, G.G. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase. Biochemistry 38, 5346–5354.
[19] Uitdehaag, J.C., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G. and Dijkstra, B.W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. Nat. Struct. Biol. 6, 432–436.
[20] Sulzenbacher, G., Bignon, C., Nishimura, T., Tarling, C.A., Withers, S.G., Henrissat, B. and Bourne, Y. (2004) Crystal structure of *Thermotoga maritima* α-L-fucosidase: insights into catalytic mechanism and the molecular basis of fucosidosis. J. Biol. Chem. 279, 13119–13128.
[21] Numao, S., Kuntz, D.A., Withers, S.G. and Rose, D.R. (2003) Insights into the mechanism of *Drosophila melanogaster* Golgi α-mannosidase II through the structural analysis of covalent reaction intermediates. J. Biol. Chem. 278, 48074–48083.
[22] Cho, S.W., Lee, S. and Shin, W. (2001) The X-ray structure of *Aspergillus aculeatus* polygalacturonase and a modeled structure of the polygalacturonase–octagalacturonate complex. J. Mol. Biol. 314, 863–878.
[23] Yip, V.L.Y., Varrot, A., Davies, G.J., Rajan, S.S., Yang, X., Thompson, J., Anderson, W.F. and Withers, S.G. (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from *Thermotoga maritima*. J. Am. Chem. Soc. 126, 8354–8355.
[24] Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R. and Huber, R. (1997) Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. J. Mol. Biol. 267, 865–880.
[25] Baxa, U., Steinbacher, S., Miller, S., Weintroub, A., Huber, R. and Seckler, R. (1996) Interactions of phage P22 tails with their cellular receptor, *Salmonella* O-antigen polysaccharide. Biophys. J. 71, 2040–2048.
[26] Deslongchamps, P. (1993) Intramolecular strategies and stereoelectronic effects. Glycosides hydrolysis revisited. Pure Appl. Chem. 65, 1161–1178.
[27] Deslongchamps, P. (1983) Stereoelectronic Effects in Organic Chemistry, Pergamon Press, Oxford, New York.
[28] Kirby, A.J. (1983) The Anomeric Effect and Related Stereoelectronic Effects at Oxygen, Springer-Verlag, Berlin.
[29] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E. (2002) The Pfam protein families database. Nucleic Acids Res. 30, 276–280. Available from: http://www.sanger.ac.uk/Software/Pfam/.
[30] Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. Available from: http://www.ebi.ac.uk/clustalw.
[31] White, A., Tull, D., Johns, K., Withers, S.G. and Rose, D.R. (1996) Crystallographic observation of a covalent intermediate in a β-glycosidase. Nat. Struct. Biol. 3, 149–154.
[32] Notenboom, V., Birsan, C., Nitz, M., Rose, D.R., Warren, R.A.J. and Withers, S.G. (1998) Insights into transition state stabilization of the beta-1,4-glycosidase Cex from *Cellulomonas fimi* by covalent intermediate accumulation in active site mutants. Nat. Struct. Biol. 5, 812–818.
[33] Nerinckx, W., Desmet, T. and Claeyssens, M. (2003) A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of *all* glycoside hydrolases. FEBS Lett. 538, 1–7.
[34] Guex, N. and Peitsch, M.C. (1997) Swiss-Model and the Swiss-PDB Viewer: an Environment for Comparative Protein Modeling. Electrophoresis 18, 2714–2723. Available from:http://www.expasy.org/spdbv/.

[35] DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA. Available from: http://www.pymol.org.
[36] Sulzenbacher, G., Driguez, H., Henrissat, B., Shülein, M. and Davies, G.J. (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolysable substrate analogue: substrate distortion gives rise to the preferred axial orientation of the leaving group. Biochemistry 35, 15280–15287.
[37] Piens, K., Ståhlberg, J., Nerinckx, W., Teeri, T.T. and Claeyssens, M. (2004) Structure-reactivity studies of *Trichoderma reesei* cellobiohydrolase Cel7A in: Lignocellulose Biodegradation (Saha, B.C. and Hayashi, K., Eds.), ACS Symposium Series, vol. 889, pp. 207–226, Oxford University Press, USA.
[38] Sulzenbacher, G., Mackenzie, L.F., Wilson, K.S., Withers, S.G., Dupont, C. and Davies, G.J. (1999) The crystal structure of a 2-fluorocellotriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 Å resolution. Biochemistry 38, 4826–4833.
[39] Sabini, E., Wilson, K.S., Danielsen, S., Shülein, M. and Davies, G.J. (2001) Oligosaccharide binding to family 11 xylanases: both covalent intermediate and mutant product complexes display $^{2,5}B$ conformations at the active centre. Acta Crystallogr., Sect. D 57, 1344–1347.
[40] Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J. and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. Protein Sci. 3, 467–475.
[41] Varrot, A. and Davies, G.J. (2003) Direct experimental observation of the hydrogen-bonding network of a glycosidase along its reaction coordinate revealed by atomic resolution analyses of endoglucanase Cel5A. Acta Crystallogr., Sect. D 59, 447–452.
[42] Hilge, M., Gloor, S.M., Rypnieuwski, W., Sauer, O., Heightman, T.D., Zimmerman, W., Winterhalter, K. and Piontek, K. (1998) High-resolution native and complex structures of thermostable β-mannanase from *Thermomonospora fusca* – substrate specificity in glycosyl hydrolase family 5. Structure 6, 1433–1444.
[43] Gebler, J.C., Trimbur, D.E., Warren, R.A.J., Aebersold, R., Namchuk, M. and Withers, S.G. (1995) Substrate-induced inactivation of a crippled β-glucosidase mutant: identification of the labeled amino acid and mutagenetic analysis of its role. Biochemistry 34, 14547–14553.
[44] Berrin, J.-G., Czjek, M., Kroon, P.A., Russell McLauchlan, W., Puigserver, A., Williamson, G. and Juge, N. (2003) Substrate (aglycone) specificity of human cytosolic β-glucosidase. Biochem. J. 373, 41–48.
[45] Ring, M., Bader, D.E. and Huber, R.E. (1988) Site-directed mutagenesis of β-galactosidase (*E. coli*) reveals that Tyr-503 is essential for activity. Biochem. Biophys. Res. Commun. 152, 1050–1055.
[46] Hogg, D., Woo, E.-J., Bolam, D.N., McKie, V.A., Gilbert, H.J. and Pickersgill, R.W. (2001) Crystal structure of mannanase 26A from *Pseudomonas cellulosa* and analysis of residues involved in substrate binding. J. Biol. Chem. 276, 31186–31192.
[47] Sakon, J., Irwin, D., Wilson, D.B. and Karplus, P.A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat. Struct. Biol. 4, 810–818.
[48] Zhou, W., Irwin, D.C., Escovar-Kousen, J. and Wilson, D.B. (2004) Mechanistic studies of *Thermobifida fusca* Cel9A mutant enzymes. Biochemistry 43, 9655–9663.
[49] Guérin, D.M.A., Lascombe, M.-B., Costabel, M., Souchon, H., Lamzin, V., Béguin, P. and Alzari, P.M. (2002) Atomic (0.94 Å) resolution structure of an inverting glycosidase in complex with substrate. J. Mol. Biol. 316, 1061–1069.
[50] Pétré, J., Longin, R. and Millet, J. (1981) Purification and properties of an endo-β-1,4-glucanase from *Clostridium thermocellum*. Biochimie 63, 629–639.
[51] Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaïch, J.-P., Driguez, H. and Hazer, R. (1998) The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. EMBO J. 17, 5551–5562.
[52] Ducros, V.M.-A., Zechel, D.L., Murshudov, G.N., Gilbert, H.J., Szabó, L., Stoll, D., Withers, S.G. and Davies, G.J. (2002) Substrate distortion by a β-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a $B_{2,5}$ conformation for the transition state. Angew. Chem. Int. Ed. 41, 2824–2827.
[53] Lovering, A.L., Lee, S.S., Kim, Y.-W., Withers, S. and Strynadka, N.C.J. (2004) Mechanistic and structural analysis of a family 31α-glycosidase and its glycosyl-enzyme intermediate. J. Biol. Chem., in press.
[54] Aleshin, A.E., Stoffer, B., Firsov, L.M., Svensson, B. and Honzatko, R.B. (1996) Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: relationship of stereochemical distortions at the nonreducing end to the catalytic mechanism. Biochemistry 35, 8319–8328.
[55] Nurizzo, D., Turkenburg, J.P., Charnock, S.J., Roberts, S.M., Dodson, E.J., McKie, V.A., Taylor, E.J., Gilbert, H.J. and Davies, G.J. (2002) *Cellvibrio japonicus* α-L-arabinanase 43A has a novel five-blade β-propeller fold. Nat. Struct. Biol. 9, 665–668.
[56] Miyanaga, A., Koseki, T., Matsuzawa, H., Wakagi, T., Shoun, H. and Fushinobu, S. (2004) Crystal structure of a family 54 α-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. J. Biol. Chem. 279, 44907–44914.
[57] Amaya, M.F., Watts, A.G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., Paris, G., Frasch, A.C., Withers, S.G. and Alzari, P.M. (2004) Structural insights into the catalytic mechanism of *Trypanosoma cruzi trans*-sialidase. Structure 12, 755–784.
[58] Watts, A.G., Damager, I., Amaya, M.L., Buschiazzo, A., Alzari, P., Frasch, A.C. and Withers, S.G. (2003) *Trypanosoma cruzi trans*-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. J. Am. Chem. Soc. 125 (25), 7532–7533.
[59] Pérez, S. and Marchessault, R.H. (1978) The exo-anomeric effect: experimental evidence from crystal structures. Carbohydrate Res. 65, 114–120.
[60] Cramer, C.J., Truhlar, D.G. and French, A.D. (1997) Exo-anomeric effects on energies and geometries of different conformations of glucose and related systems in the gas phase and aqueous solution. Carbohydrate Res. 298, 1–14.